

What is claimed is:

1 1. Apparatus for indexing a Web page which incorporates multimedia data by reference to one  
2 or more resources which supply said multimedia data, said method comprising, in combination:

3 means for analyzing said web page to identify at least one markup tag containing a  
4 reference to a given one of said resources,

5 means for selecting and executing a media processing program for analyzing the content  
6 of the multimedia data supplied by said given one of said resources to generate metadata  
7 describing said content,

8 means for formatting said metadata into a character-based text annotation,

9 means for combining said Web page and said annotation to form an enhanced Web  
10 page, and

11 means for indexing said enhanced Web page.

1 2. Apparatus as set forth in claim 1 wherein said means for selecting and executing a media  
2 processing program comprises means for determining the particular data type of the multimedia  
3 data supplied by said given resource and means for selecting a processing program for  
4 analyzing multimedia data formatted in accordance with said particular data type.

1 3. Apparatus as set forth in claim 1 wherein said means for formatting said metadata comprises  
2 means for generating a text data annotation expressed in accordance with the Extensible  
3 Markup Language.

1 4. Apparatus as set forth in claim 1 including means for acquiring additional metadata which  
2 describes the multimedia data supplied by said given one of said resources, and means for  
3 including said additional metadata in said character-based text annotation.

1 5. Apparatus as set forth in claim 4 wherein at least some of said additional data includes  
2 information obtained from said one markup tag.

1 6. Apparatus as set forth in claim 4 wherein said given resource is accessed through the  
2 operating system of a computer which provides said given resource and wherein at least some  
3 of said additional data includes information obtained from said operating system.

1 7. Apparatus as set forth in claim 4 wherein at least some of said additional information is  
2 obtained via the Internet.

1 8. Apparatus for collecting and storing metadata describing a hypertext Web page, said Web  
2 page including markup tags which identify multimedia data from one or more different external  
3 resources, said apparatus comprising, in combination,

4 a parser for identifying said markup tags in said Web page,

5 processing means for analyzing the content of said multimedia data identified by said  
6 markup tags to generate metadata describing said multimedia data,

7 means for translating said metadata into a character-based text annotation describing  
8 said multimedia data, and

9 means for storing the combination of a copy of said Web page and said annotation to  
10 form an enhanced Web page suitable for processing by text-based indexing and searching  
11 facilities.

1 9. Apparatus as set forth in claims 8 wherein said text annotation is expressed in the Extensible  
2 Markup Language.

1 10. The method of automatically enhancing the content of a Web page which contains  
2 multimedia data incorporated by reference which comprises, in combination, the steps of:  
3 identifying one or more markup tags in said Web page which respectively identify one  
4 or more external resources which provide said multimedia data;  
5 generating metadata which describes said multimedia data,  
6 translating said metadata into a character-based text annotation, and  
7 inserting said annotation into said Web page to form an enhanced Web page suitable for  
8 processing by a character-based text processing system.

1 11. The method of automatically enhancing the content of a Web page as set forth in claim 10  
2 wherein said step of identifying one or more markup tags comprises the steps of first  
3 identifying markup tags in said Web page and extracting the uniform resource locator (URL) of  
4 one of said external resources from at least selected ones of said markup tags.

1 12. The method of automatically enhancing the content of a Web page which contains  
2 multimedia data as set forth in claim 10 wherein said step of generating metadata includes the  
3 sub-steps of retrieving said multimedia data from said one or more external resources and  
4 analyzing the content of said multimedia data to extract said metadata therefrom.

1 13. The method of automatically enhancing the content of a Web page as set forth in claim 12  
2 wherein said step of generating metadata comprises the sub-steps of identifying the data type of  
3 the multimedia data from each of said resources and then selecting a processing routine for  
4 multimedia of the identified data type from each of said resources.

1 14. The method of automatically enhancing the content of a Web page as set forth in claim 10  
2 includes the further step of indexing said enhanced Web page to provide access to said Web  
3 page in response to queries expressing one or more attributes expressed in said text annotation.

1 15. The method of automatically enhancing the content of a Web page as set forth in claim 10  
2 includes the further step of searching the content of said enhanced Web page in response to a  
3 search request to determine if attributes expressed in said search request are contained in said  
4 text annotation.

1 16. The method of automatically enhancing the content of a Web page as set forth in claim 13  
2 includes the further step of indexing said enhanced Web page to provide access to said Web  
3 page in response to queries expressing one or more attributes expressed in said text annotation.

1 17. The method of automatically enhancing the content of a Web page as set forth in claim 13  
2 includes the further step of searching the content of said enhanced Web page in response to a  
3 search request to determine if attributes expressed in said search request are contained in said  
4 text annotation.